(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/31640 A1

(51) International Patent Classification[7]: G10L 21/02

(21) International Application Number: PCT/EP00/10713

(22) International Filing Date: 27 October 2000 (27.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
99203565.9 29 October 1999 (29.10.1999) EP

(71) Applicant *(for all designated States except US)*: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **HUANG, Chao-Shih,** J. [CN/NL]; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: **HOEKSTRA, Jelle**; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States *(national)*: JP, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/31640 A1

(54) Title: ELIMINATION OF NOISE FROM A SPEECH SIGNAL

(57) Abstract: A method for reducing noise in a noisy time-varying speech input signal y includes receiving the input signal y and deriving a plurality of spectral component signals representing respective magnitudes $|Y(k)|$ of spectral components of the input signal y. A correlation coefficient γ_{sn} is obtained which indicates a correlation in the spectral domain between a clean speech signal component s and a noise signal component n present in the input signal $y(y = s + n)$. Magnitudes of respective noise-suppressed spectral components $\hat{S}(k)$ are estimated by solving a correlation equation which gives a relationship between the magnitudes of the respective spectral components $|Y(k)|$ of the noisy input signal y, the spectral components $|S(k)|$ of the clean speech signal s, and the spectral components $|N(k)|$ of the noise signal n, where the equation includes the correlation based on the obtained correlation coefficient γ_{sn}. Preferably, the correlation equation is given by $|Y(k)|^a = |S(k)|^a + |N(k)|^a + \gamma_{sn}|S(k)||N(k)|$.

Elimination of noise from a speech signal.

The invention relates to a method for reducing noise in a noisy time-varying input signal, such as a speech signal. The invention further relates to an apparatus for reducing noise in a noisy time-varying input signal.

The presence of noise in a time-varying input signal hinders the accuracy and quality of processing the signal. This is particularly the case for processing a speech signal, such as for instance occurs when a speech signal is encoded. The presence of noise is even more destructive if the signal is ultimately not presented to a user, who can relatively well cope with the presence of noise, but if the signal is ultimately processed automatically, as for instance is the case with a speech signal that is recognized automatically. Increasingly automatic speech recognition and coding systems are used. Although the performance of such systems is continuously improving, it is desired that the accuracy be increased further, particularly in adverse environments, such as having a low signal-to-noise ratio (SNR) or a low bandwidth signal. Normally, speech recognition systems compare a representation of an input speech signal against a model Λ_x of reference signals, such as hidden Markov models (HMMs) built from representations of a training speech signal. The representations are usually observation vectors with LPC or cepstral components.

In practice a mismatch exists between the conditions under which the reference signals (and thus the models) were obtained and the input signal conditions. The reference signals are usually relatively clean (high SNR, high bandwidth), whereas the input signal during actual use is distorted (lower SNR, and/or lower bandwidth). It is, therefore, desired to eliminate at least part of the noise present in the input signal in order to obtain a noise-suppressed signal.

A conventional way of estimating a noise-suppressed speech signal ('clean' speech) is to use a spectral subtraction technique. In the discrete-time domain, noise speech y can be represented as:

$$y(i) = s(i) + n(i),\ 0 \leq i \leq T-1, \qquad (1)$$

where s, n, y denote clean speech, noise and noisy speech respectively, and where T denotes the length of the speech and i is the time index. The conventional spectral subtraction

technique involves determining the spectral components of the noisy-speech and estimating the spectral components of the noise. The spectral components may, for instance, be calculated using a Fast Fourier transform (FFT). The noise spectral components may be estimated once from a part of a signal with predominantly representative noise. Preferably, the noise is estimated 'on-the-fly', for instance each time a 'silent' part is detected in the input signal with no significant amount of speech signal. In the general spectral subtraction technique, the noise-suppressed speech is estimated by subtracting an average noise spectrum from the noisy speech spectrum:

$$\hat{S}(w;m) = \left[|Y(w;m)|^a - |N(w;m)|^a \right]^{1/a} e^{j\varphi_y(w;m)} \qquad (2)$$

where $\hat{S}(w;m)$, $Y(w;m)$, and $N(w;m)$ are the magnitude spectrums of the estimated speech s, noisy speech y and noise n, w and m are the frequency and time indices, respectively. The case of $a = 2$ is referred to as power spectral subtraction. The subtraction is usually called magnitude spectral subtraction if $a = 1$.

Due to the subtraction, the estimated spectrum is not guaranteed to be positive in the conventional spectral subtraction techniques. US 5,749,068 describes setting those spectral components to zero for which the subtraction yields a negative outcome:

$$\hat{S}(w) = max\{Y(w) - \alpha.N(w), 0\} \qquad (3)$$

Setting the spectral components to zero (or a low default value) is referred to as 'taking floor' for the negative spectral components. The parameter α, with a positive value, designates the degree of eliminating noise components. US 5,749,068 describes an advanced way of estimating the spectral components of the noise, but still the conventional spectral subtraction of equation (3) is used.

Taking floor for negative spectral components provides a major limitation of spectral subtraction techniques, introducing residual noise with musical tone artifacts into the estimated speech.

In order to investigate the limitation of the conventional spectrum subtraction techniques, the inventor has carried out an experiment for calculating the ratio of negative spectrum (i.e. the relative number of spectral components which would have a negative value). The negative spectrum ratio NSR_{con} for the conventional spectral subtraction technique is defined as follows:

$$NSR_{con} = \frac{1}{M}\sum_{k=0}^{M-1} f_{NS}\left(|Y(k)|^a - |\hat{N}(k)|^a\right) \quad (4)$$

$$f_{NS}(x) = \begin{cases} 1 & x < 0, \\ 0 & otherwise. \end{cases} \quad (5)$$

where $|Y(k)|$ is the corresponding magnitude spectrum of the testing speech y, $|\hat{N}(k)|$ is the noise spectrum estimated from a pause (non-speech segment), k denotes the k-th spectrum component and M represents the total number of spectral components over which the ratio is determined, for instance the number of spectral components in one frame or in the whole testing utterance.

The following table gives the negative spectrum ratio NSR_{con} for various signal to noise ratios (SNRs) with a =2. It has been found that the negative spectrum ratio NSR_{con} even reaches 34.6% at clean conditions. This illustrates that particularly at higher SNR level the conventional spectral subtraction technique introduces some residual noise, limiting the use of the technique.

SNR (dB)	Negative Spectrum Ratio (NSR_{con}) (%)
Clean	34.6
40	22.4
35	18.7
30	14.6
25	10.7
20	7.3
15	4.5
10	2.4
5	1.0
0	0.2

It is an object of the invention to overcome the limitation of the conventional spectral subtraction technique.

To meet the object of the invention, the method for reducing noise in a noisy time-varying input signal y, such as a speech signal, includes:

receiving the noisy time-varying input signal;

deriving from the signal a plurality of spectral component signals representing respective magnitudes of spectral components of the input signal;

obtaining a correlation coefficient γ_{sn} indicative of a correlation in the spectral domain between a clean speech signal component s and a noise signal component n present in the input signal ($y = s + n$); and

estimating magnitudes of respective noise-suppressed spectral components $\hat{S}(k)$ by solving an equation giving a relationship between the magnitudes of the respective spectral components $|Y(k)|$ of the noisy input signal y, the spectral components $|S(k)|$ of the clean speech signal s, and the spectral components $|N(k)|$ of the noise signal n, where the equation includes the correlation based on the obtained correlation coefficient γ_{sn}. Preferably, the correlation equation is given by:

$$|Y(k)|^a = |S(k)|^a + |N(k)|^a + \gamma_{sn}|S(k)||N(k)|$$

where a could be 1 or 2 for magnitude and power spectrum, respectively. Instead of a conventional spectral subtraction this equation is solved which is based on a correlation coefficient γ_{sn} between the clean speech s and the noise n in the spectral domain. Solving the equation can be seen as 'correlated spectral subtraction' (CSS).

The correlation coefficient γ_{sn} may be fixed, for instance based on analyzing representative input signals. Preferably, the correlation coefficient γ_{sn} is estimated based on the actual input signal. Advantageously, the estimation is based on minimizing a negative spectrum ratio. Preferably, the expected negative spectrum ratio R is defined as:

$$R = E\{f_{ns}\} = \frac{1}{M}\sum_{k=0}^{M-1} f_{ns}\left(|Y(k)|^a - |\hat{N}(k)|^a - \gamma_{sn}|\hat{S}(k)||\hat{N}(k)|\right)$$

where advantageously the 'zero-one' function f_{ns} is given by the differentiable function:

$$f_{ns}(x) = \frac{1}{1+\exp(-\alpha \cdot x + \beta)}$$

By applying the theory of adaptive learning algorithm, the correlation coefficient is advantageously obtained by following gradient operation.

$$\gamma_{sn}^{(m+1)} = \gamma_{sn}^{(m)} - \delta \nabla R$$

The correlation coefficient can be learned along the direction of NSR decrement. Preferably, this is done in an iterative algorithm.

The equation representing the correlated spectral subtraction may be solved directly. Preferably, the equation is solved in an iterative manner, improving the estimate of the clean speech.

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments shown in the drawings.

The figure shows a block diagram of a conventional speech processing system wherein the invention can be used.

General description of a speech recognition system

The noise reduction according to the invention is particularly useful for processing noisy speech signals, such as coding such a signal or automatically recognizing such a signal. Here a general description of a speech recognition system is given. A person skilled in the art can equally well apply the noise elimination technique in a speech coding system.

Speech recognition systems, such as large vocabulary continuous speech recognition systems, typically use a collection of recognition models to recognize an input pattern. For instance, an acoustic model and a vocabulary may be used to recognize words and a language model may be used to improve the basic recognition result. The figure illustrates a typical structure of a large vocabulary continuous speech recognition system 100. The following definitions are used for describing the system and recognition method:

Λ_x:	a set of trained speech models
X:	the original speech which matches the model, Λ_x
Y:	the testing speech
Λ_y:	the matched models for testing environment
W:	the word sequence
S:	the decoded sequences that can be words, syllables, sub-word units, states or mixture components, or other suitable representations.

The system 100 comprises a spectral analysis subsystem 110 and a unit matching subsystem 120. In the spectral analysis subsystem 110 the speech input signal (SIS) is spectrally and/or temporally analyzed to calculate a representative vector of features (observation vector, OV). Typically, the speech signal is digitized (e.g. sampled at a rate of 6.67 kHz.) and pre-processed, for instance by applying pre-emphasis. Consecutive samples are

grouped (blocked) into frames, corresponding to, for instance, 32 msec. of speech signal. Successive frames partially overlap, for instance, 16 msec. Often the Linear Predictive Coding (LPC) spectral analysis method is used to calculate for each frame a representative vector of features (observation vector). The feature vector may, for instance, have 24, 32 or 63 components. The standard approach to large vocabulary continuous speech recognition is to assume a probabilistic model of speech production, whereby a specified word sequence $W = w_1w_2w_3...w_q$ produces a sequence of acoustic observation vectors $Y = y_1y_2y_3...y_T$. The recognition error can be statistically minimized by determining the sequence of words $w_1w_2w_3...w_q$ which most probably caused the observed sequence of observation vectors $y_1y_2y_3...y_T$ (over time t=1,..., T), where the observation vectors are the outcome of the spectral analysis subsystem 110. This results in determining the maximum a posteriori probability:

$$\max P(W|Y, \Lambda_x), \text{ for all possible word sequences } W$$

By applying Bayes' theorem on conditional probabilities, $P(W|Y, \Lambda_x)$ is given by:

$$P(W \mid Y,\Lambda_x) = \frac{P(Y \mid W,\Lambda_x).P(W)}{P(Y)}$$

Since P(Y) is independent of W, the most probable word sequence is given by:

$$\hat{W} = \arg\max_W P(Y,W \mid \Lambda_x) = \arg\max_W P(Y \mid W,\Lambda_x).P(W) \qquad \text{(a)}$$

In the unit matching subsystem 120, an acoustic model provides the first term of equation (a). The acoustic model is used to estimate the probability P(Y|W) of a sequence of observation vectors Y for a given word string W. For a large vocabulary system, this is usually performed by matching the observation vectors against an inventory of speech recognition units. A speech recognition unit is represented by a sequence of acoustic references. Various forms of speech recognition units may be used. As an example, a whole word or even a group of words may be represented by one speech recognition unit. A word model (WM) provides for each word of a given vocabulary a transcription in a sequence of acoustic references. In most small vocabulary speech recognition systems, a whole word is represented by a speech recognition unit, in which case a direct relationship exists between the word model and the speech recognition unit. In other small vocabulary systems, for instance used for recognizing a relatively large number of words (e.g. several hundreds), or in large vocabulary systems, use can be made of linguistically based sub-word units, such as phones, diphones or syllables, as well as derivative units, such as fenenes and fenones. For such systems, a word model is given by a lexicon 134, describing the sequence of sub-word units relating to a word of the vocabulary, and the sub-word models 132, describing sequences of

acoustic references of the involved speech recognition unit. A word model composer 136 composes the word model based on the sub-word model 132 and the lexicon 134. The (sub-)word models are typically based in Hidden Markov Models (HMMs), which are widely used to stochastically model speech signals. Using such an approach, each recognition unit (word model or sub-word model) is typically characterized by an HMM, whose parameters are estimated from a training set of data. For large vocabulary speech recognition systems usually a limited set of, for instance 40, sub-word units is used, since it would require a lot of training data to adequately train an HMM for larger units. An HMM state corresponds to an acoustic reference. Various techniques are known for modeling a reference, including discrete or continuous probability densities. Each sequence of acoustic references which relate to one specific utterance is also referred as an acoustic transcription of the utterance. It will be appreciated that if other recognition techniques than HMMs are used, details of the acoustic transcription will be different.

A word level matching system 130 of The figure matches the observation vectors against all sequences of speech recognition units and provides the likelihoods of a match between the vector and a sequence. If sub-word units are used, constraints can be placed on the matching by using the lexicon 134 to limit the possible sequence of sub-word units to sequences in the lexicon 134. This reduces the outcome to possible sequences of words.

Furthermore a sentence level matching system 140 may be used which, based on a language model (LM), places further constraints on the matching so that the paths investigated are those corresponding to word sequences which are proper sequences as specified by the language model. As such the language model provides the second term P(W) of equation (a). Combining the results of the acoustic model with those of the language model, results in an outcome of the unit matching subsystem 120 which is a recognized sentence (RS) 152. The language model used in pattern recognition may include syntactical and/or semantical constraints 142 of the language and the recognition task. A language model based on syntactical constraints is usually referred to as a grammar 144. The grammar 144 used by the language model provides the probability of a word sequence $W = w_1w_2w_3...w_q$, which in principle is given by:

$$P(W) = P(w_1)P(w_2|w_1).P(w_3|w_1w_2)...P(w_q|\ w_1w_2w_3...w_q).$$

Since in practice it is infeasible to reliably estimate the conditional word probabilities for all words and all sequence lengths in a given language, N-gram word models are widely used. In an N-gram model, the term P(wj| w1w2w3...wj-1) is approximated by P(wj| wj-N+1...wj-1). In

practice, bigrams or trigrams are used. In a trigram, the term P(wj| w1w2w3...wj-1) is approximated by P(wj| wj-2wj-1).

The speech processing system according to the invention may be implemented using conventional hardware. For instance, a speech recognition system may be implemented on a computer, such as a PC, where the speech input is received via a microphone and digitized by a conventional audio interface card. All additional processing takes place in the form of software procedures executed by the CPU. In particular, the speech may be received via a telephone connection, e.g. using a conventional modem in the computer. The speech processing may also be performed using dedicated hardware, e.g. built around a DSP.

The noise elimination according to the invention may be performed in a pre-processing step before the spectral analysis subsystem 100. Preferably, the noise elimination is integrated in the spectral analysis subsystem 100, for instance to avoid that several conversions from the time domain to the spectral domain and vice versa are required. All hardware and processing capabilities for performing the invention are normally present in a speech recognition or speech coding system. The noise elimination technique according to the invention is normally executed on a processor, such as a DSP or microprocessor of a personal computer, under control of a suitable program. Programming the elementary functions of the noise elimination technique, such as performing a conversion from the time domain to the spectral domain, falls well within the range of a skilled person.

Detailed description of the invention

Details are given for speech signals. Other signals can be processed in a corresponding way. As described above, in the discrete-time domain noise speech y can be represented as:

$$y(i) = s(i) + n(i), \qquad 0 \leq i \leq T-1, \qquad (1)$$

where s, n, y denote clean speech, noise and noisy speech respectively, and where T denotes the length of the speech and i is the time index. Using conventional techniques, such as a Fast Fourier transform, the speech signal y can be transformed into a set of spectral components $Y(k)$. It will be appreciated that if already a suitable conversion to the time domain had taken place, it is sufficient to retrieve the spectral components resulting from such a conversion.

Let $|S(k)|$, $|N(k)|$, and $|Y(k)|$ be the corresponding magnitude of the spectrums of the time-domain signals s, n, and y, respectively. Using the conventional spectral subtraction techniques, individual spectral components are forced to be positive. It does not

allow the situation wherein an individual spectral component $Y(k)$ of the noisy speech y is less than the corresponding spectral component $N(k)$ of the noise signal n.

The following correlation is assumed to exist between the speech signal and the noise signal:

$$|Y(k)|^a = |S(k)|^a + |N(k)|^a + \gamma_{sn}|S(k)||N(k)| \quad (6)$$

where γ_{sn} denotes the correlation coefficient of speech and noise in the spectral domain and a could be 1 or 2 for magnitude and power spectrum, respectively. Using this correlation as the basis for estimating the clean speech spectrum (and as such using a correlated spectral subtraction) makes it possible to have the situation wherein $|Y(k)|^a < |N(k)|^a$ if $\gamma_{sn} < 0$.

Let $|\hat{S}(k)|$ and $|\hat{N}(k)|$ be the estimates of the magnitude spectrums of the clean speech signal s and the noise signal n, respectively. Preferably, $|\hat{N}(k)|$ is estimated from pause (non-speech segment). Based on equation (6), $|\hat{S}(k)|$ can be calculated by solving the equation in one step or by using an iterative algorithm. The one-step solution are give in the following equations (7) and (8) for the cases wherein a=1 or a=2, respectively:

$$|\hat{S}(k)| = \frac{|Y(k)| - |N(k)|}{(1 + \gamma_{sn}|N(k)|)}, \text{ if } a=1 \quad (7)$$

$$|\hat{S}(k)| = \frac{-\gamma_{sn}|N(k)| \pm \sqrt{\gamma_{sn}^2|N(k)|^2 + 4\left(|Y(k)|^2 - |N(k)|^2\right)}}{2}, \text{ if } a=2 \quad (8)$$

Equation (8) has two possible solutions. The positive solution which is greater than $(|Y(k)|^2 - |N(k)|^2)$ or close to $(|Y(k)|^2 - |N(k)|^2)$ will be chosen since the direction of NSR decrement is preferred.

A preferred iterative algorithm for estimating $|\hat{S}(k)|$ with specified correlation coefficient, γ_{sn}, is as follows:

LOOP k (0 : N-1)

Initialization: $\left|\hat{S}^{(0)}(k)\right|^a = |Y(k)|^a - \left|\hat{N}(k)\right|^a$ (9)

LOOP ℓ

$$|S'(k)|^a = |Y(k)|^a - \left|\hat{N}(k)\right|^a - 2\gamma_{sn}\left|\hat{S}^{(\ell)}(k)\right|\left|\hat{N}(k)\right| \quad (10)$$

$$\left|\hat{S}^{(\ell+1)}(k)\right|^a = \frac{|S'(k)|^a + \left|\hat{S}^{(\ell)}(k)\right|^a}{2} \quad (11)$$

IF $\left|\frac{\left|\hat{S}^{(\ell+1)}(k)\right|^a - \left|\hat{S}^{(\ell)}(k)\right|^a}{\left|\hat{S}^{(\ell)}(k)\right|^a}\right| <$ Threshold THEN STOP

ELSE $\ell = \ell + 1$

END LOOP ℓ

END LOOP k

The outer loop k deals with all individual spectral components. The inner loop is performed until the iteration has converged (no significant change occurs anymore in the estimated speech).

The above described algorithm can be used for a fixed correlation coefficient γ_{sn}. In a further embodiment according to the invention, the correlation coefficient γ_{sn} is estimated based on the actual input signal y. To this end, the function of negative spectrum ratio (NSR) for the correlated spectral subtraction algorithm according to the invention is defined as follows:

$$NSR = \frac{1}{M}\sum_{k=0}^{M-1} f_{NS}\left(|Y(k)|^a - \left|\hat{N}(k)\right|^a - \gamma_{sn}\left|\hat{S}(k)\right|\left|\hat{N}(k)\right|\right) \quad (12)$$

The f_{NS} function shown in equation (5) is a zero-one function. In order to derive the relation between the correlation coefficient γ_{sn} and NSR, a smoothed zero-one, sigmoid function family is preferably used. For example, the following function f_{ns} is advantageously used for further derivation due to its differentiability.

$$f_{ns}(x) = \frac{1}{1+\exp(-\alpha \cdot x + \beta)} \quad (13)$$

Exemplary values for α and β are 1.0 and 0.0, respectively.

Then, the expected negative spectrum ratio R is defined as follows:

$$R = E\{f_{ns}\} = \frac{1}{M}\sum_{k=0}^{M-1} f_{ns}\left(|Y(k)|^a - |\hat{N}(k)|^a - \gamma_{sn}|\hat{S}(k)||\hat{N}(k)|\right) \quad (14)$$

By applying the theory of adaptive learning algorithm, the correlation coefficient is preferably obtained by the following gradient operation:

$$\gamma_{sn}^{(m+1)} = \gamma_{sn}^{(m)} - \delta\nabla R \quad (15)$$

The correlation coefficient can be learned along the direction of decrease in NSR. This implies to reduce the residual noise in the estimated spectrum using the proposed correlated spectral subtraction (CSS) algorithm.

The algorithm of estimating $|\hat{S}(k)|$ with a minimum NSR based correlation coefficient γ_{sn} is as follows:

Initialization: $m = 0$

$\gamma_{sn}^{(m)}$ = non-zero, initial guess.

LOOP m

Block 1

LOOP k (0 : N-1)

$\ell = 0$

$|\hat{S}^{(\ell)}(k)|^a = |Y(k)|^a - |\hat{N}(k)|^a$

LOOP ℓ

$$|S'(k)|^a = |Y(k)|^a - |\hat{N}(k)|^a - \gamma_{sn}^{(m)}|\hat{S}^{(\ell)}(k)||\hat{N}(k)|$$

$$|\hat{S}^{(\ell+1)}(k)|^a = \frac{|S'(k)|^a + |\hat{S}^{(\ell)}(k)|^a}{2}$$

$$\text{If } \frac{\left||\hat{S}^{(\ell+1)}(k)|^a - |\hat{S}^{(\ell)}(k)|^a\right|}{|\hat{S}^{(\ell)}(k)|^a} < \textit{Threshold 1} \text{ THEN STOP}$$

ELSE $\ell = \ell + 1$

END LOOP ℓ

END LOOP k

$$R = E\{f_{ns}\} = \frac{1}{M}\sum_{k=0}^{M-1} f_{ns}\left(|Y(k)|^a - |\hat{N}(k)|^a - \gamma_{sn}|\hat{S}(k)||\hat{N}(k)|\right)$$

$$\gamma_{sn}^{(m+1)} = \gamma_{sn}^{(m)} - \delta\nabla R$$

If $\left|\frac{\gamma_{sn}^{(m+1)} - \gamma_{sn}^{(m)}}{\gamma_{sn}^{(m)}}\right| < Threshold\ 2$ THEN STOP

END LOOP m

The block indicated as block 1 is the same as used for the iterative algorithm assuming a fixed correlation coefficient γ_{sn}. Instead of using the iterative solution in block one, also the one-step solution of equations (7) or (8) may be used.

It will be appreciated that after the noise has been eliminated as described above, the resulting estimated spectral components of the noise-eliminated signal may be converted back to the time-domain. Where possible the spectral components may be used directly for the subsequent further processing, like coding or automatically recognizing the signal.

CLAIMS:

1. A method for reducing noise in a noisy time-varying input signal y, such as a speech signal; the method including:

receiving the noisy time-varying input signal y;

deriving from the input signal y a plurality of spectral component signals representing respective magnitudes $|Y(k)|$ of spectral components of the input signal y;

obtaining a correlation coefficient γ_{sn} indicative of a correlation in the spectral domain between a clean speech signal component s and a noise signal component n present in the input signal y ($y = s + n$); and

estimating magnitudes of respective noise-suppressed spectral components $\hat{S}(k)$ by solving a correlation equation giving a relationship between the magnitudes of the respective spectral components $|Y(k)|$ of the noisy input signal y, the spectral components $|S(k)|$ of the clean speech signal s, and the spectral components $|N(k)|$ of the noise signal n, where the equation includes the correlation based on the obtained correlation coefficient γ_{sn}.

2. The method as claimed in claim 1, wherein the correlation coefficient γ_{sn} is predetermined.

3. The method as claimed in claim 1, wherein the step of obtaining the correlation coefficient γ_{sn} includes estimating the correlation coefficient γ_{sn}.

4. The method as claimed in claim 3, wherein the step of estimating the correlation coefficient γ_{sn} includes determining a minimum negative spectrum ratio.

5. The method as claimed in claim 4, wherein the negative spectrum ratio NSR represents a proportion of spectral components $\hat{S}(k)$ which would be negative based on the solution of the correlation equation.

6. The method as claimed in claim 5, wherein the method includes:

initializing the correlation coefficient γ_{sn} with a non-zero value; and

iteratively:

performing the step of solving the correlation equation to obtain $|\hat{S}(k)|$;

and

estimating a new correlation coefficient based on a gradient decent of the negative spectrum ratio NSR for $\hat{S}(k)$.

7. The method as claimed in claim 1, wherein the step of solving the correlation equation includes iteratively estimating the noise-suppressed spectrum $\hat{S}(k)$.

8. The method as claimed in claim 7, wherein method includes calculating an initial estimate of a magnitude of the noise-suppressed spectrum $\hat{S}^{(0)}(k)$ by subtracting a magnitude of an estimate of the respective spectral components $\hat{N}(k)$ of the noise signal n from a magnitude of the respective spectral components $Y(k)$ of the noisy input signal y.

9. The method as claimed in claim 7, wherein the step of performing the iterative spectrum estimation includes in each iteration:

estimating a magnitude of an auxiliary noise-suppressed spectrum based on the correlation equation where a term with the correlation coefficient γ_{sn} is based on a current estimate of a magnitude of the noise-suppressed spectrum $\hat{S}^{(l)}(k)$; and

estimating a new magnitude of the noise-suppressed spectrum $\hat{S}^{(l+1)}(k)$ based the estimated magnitude of the auxiliary noise-suppressed spectrum and on the current estimate of a magnitude of the noise-suppressed spectrum $\hat{S}^{(l)}(k)$.

10. An apparatus for reducing noise in a noisy time-varying input signal y, such as a speech signal; the apparatus including:

an input for receiving the noisy time-varying input signal y;

means for deriving from the input signal y a plurality of spectral component signals representing respective magnitudes $|Y(k)|$ of spectral components of the input signal y;

means for obtaining a correlation coefficient γ_{sn} indicative of a correlation in the spectral domain between a clean speech signal component s and a noise signal component n present in the input signal y ($y = s + n$); and

means for estimating magnitudes of respective noise-suppressed spectral components $\hat{S}(k)$ by solving a correlation equation giving a relationship between the magnitudes of the respective spectral components $|Y(k)|$ of the noisy input signal y, the spectral components $|S(k)|$ of the clean speech signal s, and the spectral components $|N(k)|$ of the noise signal n, where the equation includes the correlation based on the obtained correlation coefficient γ_{sn}.

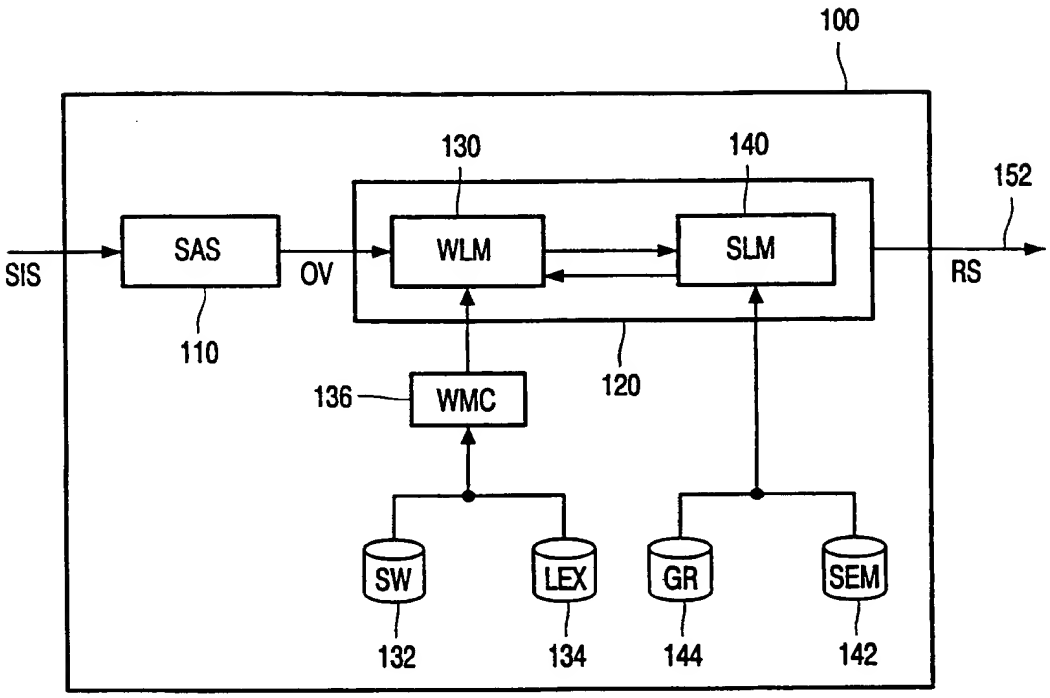
100
130
140
152
SIS
SAS
OV
WLM
SLM
RS
110
120
136
WMC
SW
LEX
GR
SEM
132
134
144
142

fig. 1

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION**	see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.
PHN 17.717WO		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/EP 00/ 10713	27/10/2000	29/10/1999
Applicant		
KONINKLIJKE PHILIPS ELECTRONICS N.V.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 2 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language,** the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

   [ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

   [ ] contained in the international application in written form.

   [ ] filed together with the international application in computer readable form.

   [ ] furnished subsequently to this Authority in written form.

   [ ] furnished subsequently to this Authority in computer readble form.

   [ ] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   [ ] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).

3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title,**

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. –

   [ ] as suggested by the applicant.

   [ ] because the applicant failed to suggest a figure.

   [ ] because this figure better characterizes the invention.

   [X] None of the figures.

INTERNATIONAL SEARCH REPORT

Int tional Application No
PCT/EP 00/10713

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L21/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HUANG J ET AL: "An energy-constrained signal subspace method for speech enhancement and recognition in white and colored noises" SPEECH COMMUNICATION,NL,ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, vol. 26, no. 3, November 1998 (1998-11), pages 165-181, XP004152155 ISSN: 0167-6393 the whole document	1-10
A	US 5 749 068 A (SUZUKI) 5 May 1998 (1998-05-05) cited in the application abstract	1-10

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 February 2001	23/02/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Quélavoine, R

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int tional Application No
PCT/EP 00/10713

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5749068 A	05-05-1998	JP 9258768 A	03-10-1997